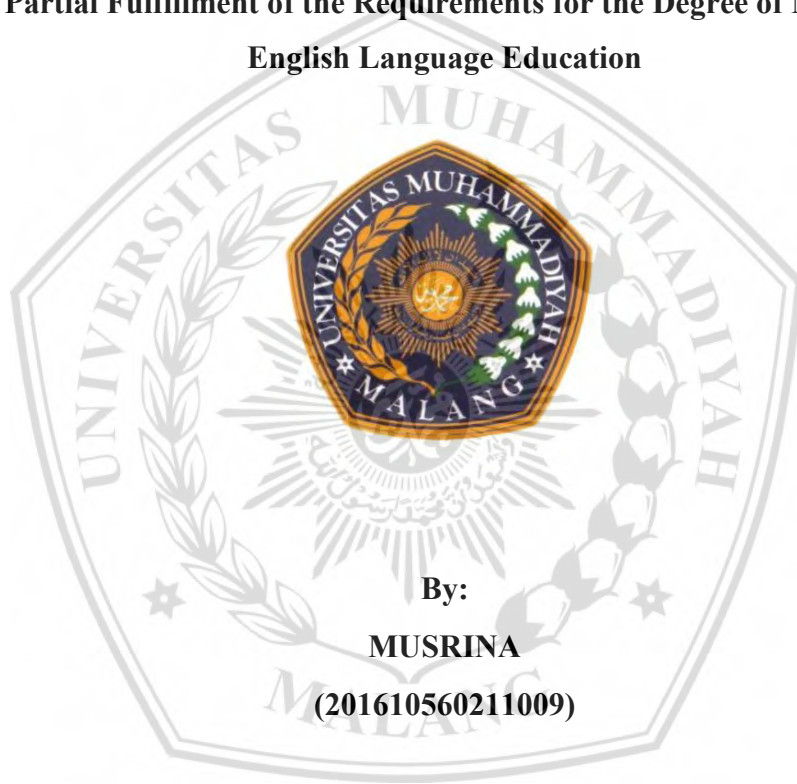


**ANALYTIC SCORING TECHNIQUE FOR PEER ASSESSMENT FOR
STUDENTS' SPEAKING SKILL AT ENGLISH DIPLOMA 3
UNIVERSITY OF MERDEKA MALANG**

THESIS

**In Partial Fulfillment of the Requirements for the Degree of Master of
English Language Education**



**By:
MUSRINA
(201610560211009)**

**DIRECTORATE OF POSTGRADUATE PROGRAM
UNIVERSITY OF MUHAMMADIYAH MALANG
SEPTEMBER 2018**

**ANALYTIC SCORING TECHNIQUE FOR PEER ASSESSMENT FOR
STUDENTS' SPEAKING SKILL AT ENGLISH DIPLOMA 3 UNIVERSITY
OF MERDEKA MALANG**

Proposed by:

**MUSRINA
201610560211009**

Has been accepted on
Friday, 7th September 2018

Advisor I



Dr. Hartono

Advisor II



Dr. Sudiran

Director of Directorate of
Postgraduate Program



Akhsanul In'am, Ph.D

Head of Program Study




Dr. Hartono

THESIS

Written by:

MUSRINA
201610560211009



Has been examined in front of examiners
On Friday, 7th September 2018 and decided that
It has fulfilled the requirements to get
Master Degree of English Language Education
in Postgraduate Program of University of Muhammadiyah Malang

The Examiners

Chief : Dr. Hartono

Secretary : Dr. Sudiran

1st Examiner : Dr. Sri Hartiningsih

2nd Examiner : Bayu Hendro Wicaksono, Ph.D

LETTER OF STATEMENT

I, the undersigned:

Name : **MUSRINA**
NIM : **2016105602110091**
Study Program : **Master of English Language Education**

Hereby, I declare that:

1. The thesis entitled: **ANALYTIC SCORING TECHNIQUE FOR PEER ASSESSMENT FOR STUDENTS' SPEAKING SKILL AT ENGLISH DIPLOMA 3 UNIVERSITY OF MERDEKA MALANG** is my original work and contains ~~no~~ one's scientific paper that may be proposed to achieve an academic degree at ~~any~~ universities. Besides, there is no other's idea or citation except those which have been quoted and mentioned at the bibliography.
2. If this thesis is proven as a form of **PLAGIARISM** in this thesis, I am willing to accept the consequences including accepting the **CANCELLATION OF THE GRANTING OF MASTER DEGREE** and undergoing any procedures required by the prevailing law.
3. This thesis can be used for literature review which can be accessed by others freely (**NON EXCLUSIVE ROYALTY**).

Thus, this statement is made truthfully to be used as appropriate.

Malang, 7th September 2018

The Writer,



MUSRINA

MOTTO AND DEDICATION

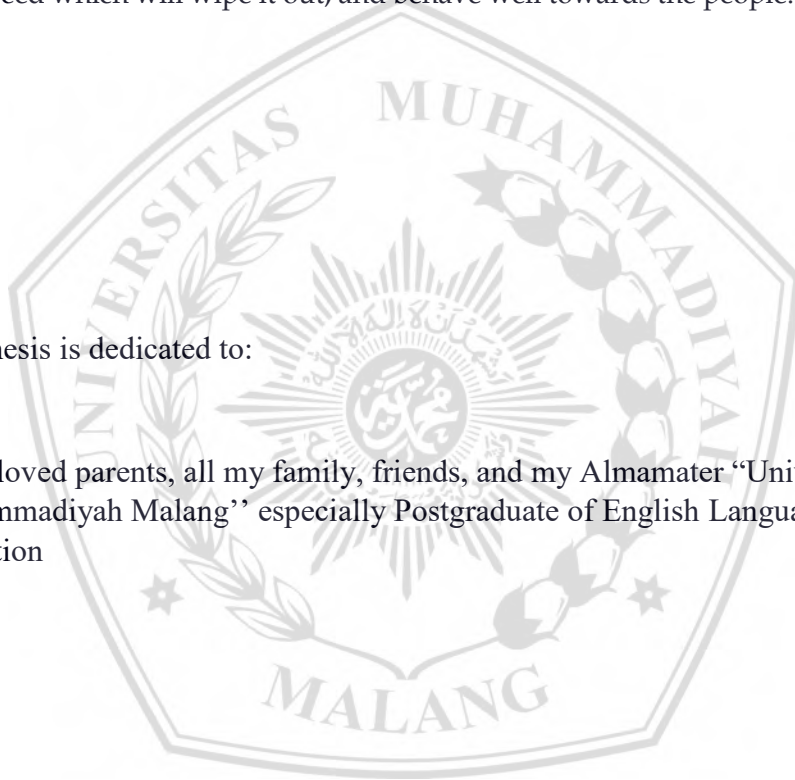
MOTTO

يَتَّقِ اللَّهَ حَيْثُمَا لُتُّ، وَيَتَّبِعِ لِرِوَاةٍ لِحَدِيثٍ قَدْ مَحَدَا، وَخَلَقِ لِنَاسٍ عُنُقٍ حَنَنَ (Al hadist)

Have taqwa (fear) of Allah wherever you may be, and follow up a bad deed with a good deed which will wipe it out, and behave well towards the people.

This thesis is dedicated to:

My beloved parents, all my family, friends, and my Almamater “University of Muhammadiyah Malang” especially Postgraduate of English Language Education



ACKNOWLEDGMENTS

Alhamdulillah, praise to Allah almighty, the most merciful and the most compassionate who has blessed the researcher with substantially uncountable blessing. One of those, Allah is the turner of the hearts who has turned the researcher's heart to be strongly and continually committed to finish the research rapidly. Praise and invocation are also given to our Great Prophet, Muhammad (peace be upon him) as the most influential prophet who had given us the real example of how to always stick to Allah during the ups and downs.

Furthermore, at this section, the researcher would like to express her best regard and gratitude to her first advisor Dr. Hartono, M.Pd and her second advisor Dr. Sudiran, M.Hum. The advisors, they had spared their busy time to have detail check and recheck the researcher's piece of work until can be worthy presented as the one of requirement of getting the master degree. For their effort and their kindness throughout the guidance, there is nothing much the researcher can do but pray for them to have long and prosperous life.

At this moment the researcher also likes to thank the English Diploma 3 of University of Merdeka Malang. This institution was very supporting and gave the warm welcome to the researcher to conduct the experimental research in their classes. Moreover, the students also were very enthusias and supportif during the research process.

The last but not least, the researcher would like to thank her wonderful family and friends. For the family, the researcher is beyond grateful to have a father and a mother who always give the the valuable advise and never stop praying for the researcher's success. For friends, thank you for the help, the support and wonderful togetherness that shared throughout the years.

Malang, 5 September 2018

TABLE OF CONTENTS

APPROVAL.....	ii
LETTER OF STATEMENT.....	iii
MOTTO AND DEDICATION.....	iv
ACKNOWLEDGMENTS	v
TABLE OF CONTENTS.....	vi
ABSTRACT	viii
ABSTRAK	ix
INTRODUCTION.....	1
LITERATURE REVIEW.....	4
Scoring Technique for Speaking Skill	4
Types of Scoring Technique for Speaking Skill	5
Holistic Scoring Technique	5
Demerit of Holistic Scoring Technique.....	7
Analytic Scoring Technique.....	8
Merit of Analytic Scoring Technique.....	8
Demerit of Analytic Scoring Technique.....	9
Peer Assessment.....	10
The Way to Employ Peer Assessment.....	10
RESEARCH METHOD	11
Research Design.....	11
Research Variable	12
Setting and Population Research	12
Research Instruments	13
Speaking Tests.....	13
Scoring Scale	14
Rating Process	14
Test Validity	15
Reliability.....	17
Research Procedure	18
Procedure for the Experimental Group.....	18
Procedure for the Control Group.....	18

Data Analysis	19
FINDINGS AND DISCUSSION	19
Research Findings	20
Discussion	23
CONCLUSION AND SUGGESTION	25
Conclusion	25
Suggestions	26
REFERENCES	26

APPENDIXES

Appendix 1 : Speaking Test.....	31
Appendix 2 : Scoring Scale.....	33
Appendix 3 : Lesson Plan	35
Appendix 4 : Students' Attendance List.....	52
Appendix 5 : The Students Speaking Scores	56
Appendix 6 : Statistical Analysis	64

Analytic Scoring Technique for Peer Assessment for Students' Speaking Skill at English Diploma 3 University of Merdeka Malang

Musrina

Master of English Language Education, University of Muhammadiyah Malang,
musrinaazakir@gmail.com

ABSTRACT

This study was intended to know whether or not the use of analytic scoring technique for speaking peer assessment could contribute significant gains on enhancing the students' speaking skill compared to those assessed by holistic scoring technique. This research employed quasi-experimental design with non-equivalent (Pre-Test and Post-Test) control group design. The experimental research was conducted at University of Merdeka Malang specifically at English Diploma III. Two classes were chosen during the research as the experiment and the control group. Analytic scoring technique for speaking peer assessment was employed in experimental group. On the other way around, holistic scoring technique for speaking peer assessment was employed in control group. Based on the result of independent sample t-test score, it was observed that t value (3.040) was higher than t table (2.01940). Thus, it showed that the alternative hypothesis was accepted. Seeing from sig.(2 tails) the significance level of the two groups was .004 less than .05. Hence, it can be concluded that the null hypothesis was rejected, and the alternative hypothesis was accepted. In other words, the students' speaking skill of post test result in the experimental group significantly improved compared with the students' speaking result in the control group assessed by peer assessment using holistic scoring technique.

Keywords: *Analytic Scoring Technique, Holistic scoring technique, Peer Assessment, Speaking Skill*

**TEKNIK PENILAIAN ANALITIK PADA PENILAIAN SEBAYA UNTUK
MAHASISWA DIPLOMA 3 JURUSAN BAHASA INGGRIS DI
UNIVERSITAS MERDEKA MALANG**

Musrina

Magister Pendidikan Bahasa Inggris, Universitas Muhammadiyah Malang,
musrinaazakir@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk mengetahui apakah penggunaan teknik penilaian analitik dalam *Speaking* pada penilaian sejawat dapat memberikan kontribusi yang signifikan pada peningkatan keterampilan berbicara mahasiswa dibandingkan dengan mereka yang dinilai dengan teknik penilaian holistik. Penelitian ini menggunakan desain quasi-eksperimental dengan (*Pre-Test dan Post-Test*) desain kelompok kontrol. Penelitian eksperimental dilakukan di Universitas Merdeka Malang khususnya di Diploma III jurusan Bahasa Inggris. Dua kelas dipilih selama penelitian sebagai kelas eksperimen dan kelas kontrol. Teknik penilaian analitik pada penilaian sejawat diterapkan di kelompok eksperimen. Sebaliknya, teknik penilaian holistik pada penilaian sejawat diterapkan di kelompok kontrol. Berdasarkan hasil perhitungan menggunakan *independent sample t-test*, diamati bahwa nilai t (3.040) lebih tinggi dari t tabel (2,01940). Dengan demikian, itu menunjukkan bahwa hipotesis alternatif diterima. Adapun dilihat dari sig. (2 tailed) juga menunjukkan bahwa tingkat signifikansi dari dua kelompok itu 0,004 kurang dari 0,05. Dengan demikian dapat disimpulkan bahwa null hipotesis ditolak, dan alternatif hipotesis diterima. Dengan kata lain, keterampilan berbicara mahasiswa dari hasil posttest pada kelompok eksperimen secara signifikan meningkat dibandingkan dengan hasil berbicara mahasiswa pada kelompok control yang dinilai dengan penilaian sejawat menggunakan teknik penilaian holistik.

Kata kunci: *Teknik Penilaian Analitik, Teknik Penilaian Holistik, Penilaian Sejawat, Kemampuan Berbicara*

INTRODUCTION

Teaching speaking can be enormously challenging for the teachers. Teachers are required to be able to encourage students to speak by providing proper activities. A Study in Indonesian context had been conducted by Febrianti & Rosana (2011) related to the problems in teaching speaking in which she found that some students were most likely to stay silence and failed to express their idea orally. According to Brown (2004), due to an inadequate chance to get exposure target language and have interaction with native speakers, the EFL learners in general relatively poor at spoken English, especially regarding fluency, control of idiomatic expressions, and understanding of cultural pragmatics. Torky (2006) elaborates some issues that students generally encounter in learning speaking, namely the conflicts between the fluency and accuracy, lack of confidence and pronunciation problem. Therefore, Solcova (2011) suggested that English teachers teach speaking by training the learners to speak accurately (in term of, for example, pronunciation, grammatical) or arousing them to speak fluently.

Hosseini and Mowlaie (2016) claim that because enhancing students' speaking skill is challenging, the teachers should not only focus on trying to implement the appropriate teaching methods but also should employ the proper techniques of assessing students' speaking performance. Proper techniques for assessment are required to inform the teachers about the students' progress and what to develop. Hence, the teachers are required to have good understanding about language assessment specifically how to measure oral performance of the learner.

Brown (2004) states that a well-constructed test must set an accurate measure of test-takers ability, and constructing a good test is a complex task including both science and art. In this sense, the teacher must know the appropriate task for speaking and support it with the proper scoring technique. When the scoring technique is used improperly during the assessment, the result of the assessment can debilitate the learners' score which will unquestionably affect their achievement. Tuan (2012) revealed in his study that the use of inappropriate scoring technique can render the students some drawbacks to their autonomy process of learning speaking.

Theoretically speaking, there are two techniques of scoring speaking, namely the holistic (impressionistic) and the analytic scoring (Heaton, 1989; Mcnamara, 2000; Brown, 2004). Holistic scoring is assessing speaking performance where the assessor evaluates by looking the whole performance based on the impression (Fulcher, 2003). This type of scoring had emerged for a very long time and known as traditional way of scoring production skill (Lumley, 2005). Therefore, it has been very regular to English teachers. Tuan (2012) discussed in his study that the use of holistic scoring in Vietnam has been the foundation for teaching and speaking assessment. On the other hand, analytical scoring is defined as a type of scoring productive skills by reflecting several aspects of language components (Brown, 2004). This type of scoring seems complete because it concerns on linguistic components to be assessed such as grammar, pronunciation, vocabulary, and etc. Fairbairn and Dunlea (2017) argue that analytic scoring is preferable in instructional setting because the students' weaknesses and strengths are diagnosed. However, employing analytic scoring technique might be challenging for the examiner whom at same time as the interlocutor to assess multiple scores where she or he may not have extra time to focus on complex evaluation (O'Sullivan, 2008). In addition, applying analytical scoring also might be time consuming when it deals with the large class with limited time (Tuan, 2012).

In response to the challenge above, applying peer assessment can be the solution to decrease the challenge by providing an opportunity to classmates to assess other students' peer performance. Central of Academic Development (2013) states that involving peers as a strategy in teaching learning has been used for many years; yet, widespread of peer assessment is some more recent phenomena. Biggs and Tangs (2007) claim that peer assessment can promote the interpersonal skill of teamwork and reduce teachers' marking load. Peer assessment is a natural extension of the move from a teacher-centred to a student-centred mode of education, which emphasises the active engagement of students in their learning, learner responsibility, metacognitive skills and a dialogical, collaborative model of teaching and learning (Spiller, 2012).

As a result, due to several reasons in preceding paragraphs, this study was intended to investigate the effect of analytic scoring technique for speaking peer assessment on the enhancement of the students' speaking skill. Some of researchers have observed the topic about analytic scoring. Hosseini and Mowlaie (2016) studied about the effect of holistic and analytic assessment in Iranian intermediate EFL learners' writing skill. The result indicated that both holistic and analytic showed the significant difference where the group employing analytic scoring outperformed the holistic group. In speaking area, Tuan (2012) has conducted research about assessing speaking skill using analytic scoring technique in Vietnam context. The result showed that the research provides an insight into the interconnection between the analytic approach and students' improvement in speaking learning.

Although, the research has provided an insight about the advantage of analytic scoring, this type of scoring technique is still less preferable compared with holistic scoring technique. Ounis (2017) who studied about a comparison between holistic and analytic assessment of speaking found that there were many differences between the two scoring techniques. Apart from the differences, it was noted that the holistic scale was more useful, reliable and consistent as far as speaking assessment is concerned. Önem (2015) investigated teachers' attitude toward holistic and analytic scoring for assessing students' oral performance. The result revealed that EFL Turkish teachers preferred to employ holistic than analytic scoring due to its practicality. Due to the fact, this study was intended to explore the effect of analytic scoring technique for peer assessment in enhancing students' speaking performance in order to minimize the time while employing analytic scoring technique.

Preliminary research conducted in University of Merdeka Malang at 15 January 2018 showed that the final achievement of students' speaking competence during the first semester of learning was still considered low. The speaking teacher during the informal interview also argued that some teaching techniques and cutting-edge media had been employed to enhance students' speaking competence. However, most of the students still struggled to communicate in English properly.

When it came to assessment, the teacher stated that she mostly used field notes. Mukminatin (2000) who conducted a study about the advantages analytic scoring procedure found that some of speaking teachers still used their own techniques in assessing students' speaking performance which was field note not fully effective. These field notes were not effective because the teacher could possibly miss to overall diagnose the strength and weaknesses of students' performance. Therefore, this present study was aimed to seek to discern the effectiveness of analytic scoring technique for speaking peer assessment in enhancing students' speaking skill.

Based on the background of the study, the research question was formulated as follows:

“Does the use of analytic scoring technique for speaking peer assessment contribute significant gains on enhancing the students' speaking skill compared to those assessed by holistic scoring technique?”

In this study, the researcher formulated two different hypothesis as follows:

Null Hypothesis = Assessing students by using analytic scoring technique for speaking peer assessment does not contribute significant gains on enhancing the students' speaking skill compared to those assessed by holistic scoring technique.

Alternative Hypothesis = Assessing students by using analytic scoring technique for speaking peer assessment contributes significant gains on enhancing the students' speaking skill compared to those assessed by holistic scoring technique.

LITERATURE REVIEW

Scoring Technique for Speaking Skill

Fairbairn and Dunlea (2017) define scoring technique for productive skill as a process to rate the written and the spoken task which consists some components or criteria to be measured. O'Sullivan (2015b) describes the distinction of the process between measuring rating written and spoken task. For writing, raters have access to the written performance and can move easily back and forth over the response, revisiting and rechecking salient parts of the performance. In speaking,

real-time processing is required for rating and the process of revisiting sections of the performance is much more difficult and time-consuming.

Types of Scoring Technique for Speaking Skill

There are two types of techniques used to assess oral performance. According to Heaton (1989), the scoring of speaking assessment can be measured by the impression mark and detail marking scheme. This goes in line with Brown (2004) who divides two types of measuring oral performance, namely holistic technique (impression mark) and analytic technique. The detail comparison of these two techniques will be elaborated as follows.

Holistic Scoring Technique

Holistic scoring technique is also called as a traditional process of measuring productive skill. Brown (2004) elaborates that in holistic scoring technique, the teacher or rater uses an overall impression related to the performance to arrive at a score. In addition, Alderson, Clapham & Wall (1995) state that in holistic technique, “the examiners are asked not to pay too much attention to any one aspect of a candidate’s performance but rather to judge its overall effectiveness”. In a similar vein, Fulcher (2003) pointed out that the holistic technique is created to summarize all the characteristics of the sample to represent the whole quality of the performance. In other words, the teachers’ impression related to the students’ score can be obtained by reflecting to the quality of oral performance.

The Way to Employ Holistic Scoring Technique

There are several aspects which need to be considered during the implementation of holistic scoring technique. Firstly, the raters do not pay attention to one specific element but overall performance (Brown, 2004). Secondly, the rater uses holistic scoring scale to facilitate the scoring activity. In the holistic scoring scale, the criteria falls into a very good speaker, a good speaker, an average speaker, a poor speaker and a very poor speaker (Carroll, 1980). For example, the indicator of a very good speaker can be described as a speaking with authority on variety of topics who can initiate, expand and develop a theme. However, the indicators of being a very poor speaker can be elaborated as having serious problems in

communication, having limited vocabulary or her or his utterance cannot be understood by the listener. Some of the indicator criteria can be seen in Appendix 2b in holistic scoring scale. Lastly, the grading system can be reached by the impression based on overall quality of the performer (Brown, 2004).

Merit of Holistic Scoring Technique

Brown (2004) suggests several advantages of employing holistic scoring technique. Firstly, holistic technique offers fast evaluation. The use of holistic technique can minimize the time for the rater. Moreover, Jamieson and Poonpon (2013) state that holistic scoring is suitable for situations in which a large number of test takers are assessed, and scoring has to be carried out in a short time, such as for admission decisions. In this sense, the rater does not need to concentrate on separate mechanisms during evaluation. Hence, this type of scoring can be helpful to be employed in assessing speaking in large class. Heaton (1989) argues that one of the difficulties of testing oral performance is in the time limit of administration. Therefore, to maximize the time, holistic scoring technique can facilitate the teacher to have fast evaluation.

Secondly, holistic scoring technique has high inter-rater reliability (Brown, 2004). Inter-rater reliability refers to the degree of agreement among raters. For example, if the raters agree 8 out of 10 times, meaning that the test has 80 % of inter-rater reliability (Cherry, 2009). Inequality result commonly occurs due to several factors such as lack of attention to scoring criteria, inexperience, or preconceived biases (Brown, 2004). Ounis (2017) conducted a comparison study about holistic and analytic scoring techniques. The result showed that holistic scoring technique had outperformed analytic scoring technique in term of inter-rater reliability with comparison mark of 62.22% of holistic technique and 57,037% of analytic technique. However, the result of the study cannot be considered having high inter-rater reliability because the result still below 70% as the acceptable average of agreement. As a result, although holistic technique has not reached the maximum agreement of inter-rater reliability, it showed better reliability result compared to analytic scoring technique.

Lastly, holistic scoring technique represents a standard that is simply to be understood or interpreted by common raters (Brown, 2004). Huot (1990b) also states that holistic scoring technique is flexible due to its simple implementation. Fulcher (2010) points out that some of raters prefer to use holistic scoring technique because they do not have to be controlled with very detailed and restrictive sets of criteria of evaluation.

Demerit of Holistic Scoring Technique

On the other hand, holistic scoring technique has also some disadvantages. Brown (2004) states that holistic scoring technique has no diagnostic information related to the performance. In other words, by utilizing this scoring technique, the teachers who acts as the raters will barely recognize the weakness and the strength of the test taker. Therefore, the holistic technique does not provide clear feedback related to the test performance. It does not facilitate feedback because the concentration is broad. Dickins, 2004; Fulcher, 1996b; Kelly & Turner, 2007; Upshur & Turner, 1995 as cited in Jamieson and Poonpon (2013) argue that holistic scoring is too broad to apprehend speaking improvement of students and hardly facilitate the teacher with satisfactory guidance for formative classroom assessment. That issue happens due to the fact that holistic scoring technique combines all the descriptions of features that make the result hard to interpret (Fairbairn and Dunlea, 2017).

In addition, holistic scoring technique technique has low validity. Weigle (2002) states that holistic scoring is less acceptable in term of validity. As stated by Fulcher (2003), the final product is composed of different pieces and ignoring the pieces may lead to inconclusive or misleading results. For instance, when students have fluently performed oral performance but have some grammatical errors, it will lead misleading result when assessed by another scoring technique that concerns on grammatical feature. Furthermore, Charney, 1984; Cumming, 1990; Lyons, 1990; Reid, 1993; Cohen and Manion, 1994; White, 1994; Elbow, 1999 as cited in Bacha (2001) belief that the rater in holistic scoring technique focuses more what they understand well rather on the specific area of the rater's weakness. However, according to Fulcher and Davidson (2007), the rater can be trained comprehensively

to reach the validity of scoring. In other words, in order to reach the valid score during evaluation, the rater must comprehend every feature that is crucial in speaking.

Analytic Scoring Technique

Analytic scoring technique is as a type of scoring productive skills by reflecting several aspects of language components (Brown, 2004). According to McNamara (2000), analytic scoring technique separates specific features of language to arrive at a score. In a similar vein, Alderson, Clapham & Wall (1995) explain that analytic scoring technique analyze students' oral performance focuses on various mechanisms along with the descriptors given at different levels for each feature. As a result, the scoring procedure in this multi elements- based analysis is gained by combining of overall results from separate aspects. In other words, the overall score can be calculated by accumulating all results or considering and valuing the scores separately (Alderson, Clapham & Wall, 1995).

The Way to Employ Analytic Scoring Technique

The procedure to employ analytic scoring technique consists several steps. Firstly, separating the several features of a macro and micro skill of speaking for scoring purposes. For example, speaking consists of micro skill where the learners produce the small chunk of languages like (phoneme, morpheme, word, collocation and phrasal unit) and macro skill where the learners produce on large elements such as fluency, discourse, function, style, cohesion, non-verbal communication, and strategic option (Brown, 2004). Secondly, simplifying the process of scoring by combining the elements of micro and macro speaking skill and put it into analytic scoring scale. For example, the analitic scoring scale consists of pronunciation, vocabulary, grammar, and fluency. Therefore, during the scoring, the rater will focus on judging the strenghts and the weakneses of the students' speaking performance by seperating the criteria. Lasly, the scoring prosess in analitic scoring technique can be reached by accumulating the score from every aspects.

Merit of Analytic Scoring Technique

There are some benefits of analytic scoring technique. Firstly, this type of scoring is highly appropriate in the instruction used. Fairbairn and Dunlea (2017)

argue that analytic scoring technique is preferable in instructional setting because the students' weaknesses and strengths are diagnosed. The weakness and strength can easily be identified because it separates the diagnose into several features of language components. This goes in line with Tuan (2012) who states that the analytic technique will facilitate students with detail information related to their competence. In addition, due to its diagnostic aspect, analytic scoring will facilitate the teacher in giving feedback related to the students' performance (Fulcher & Davidson (2007). This goes in line with Brown (2004) who states that analytic scoring focuses on some features of a performance that can give comprehensive information about the performance. Moreover, analytic scoring technique is flexible in term of expectation. It provides a specific set of language abilities to test that can facilitate the teacher or the rater to differentiate and consider which components based on the expectations (Alderson, Clapham & Wall, 1995; Bachman & Palmer, 1996).

Demerit of Analytic Scoring Technique

Despite the advantages of analytic scoring technique, some disadvantages also can be considered. Firstly, analytic scoring technique is time consuming. According to Tuan (2012), analytical scoring needs a lot of time to rate because the raters are expected to score several aspects in single performance. This goes in line with Hughes (2003) who states that analytic needs more time than holistic scoring no matter how extensively and well the raters are trained. This type of scoring may not be appropriate for large classes. Secondly, Luoma (2004) points out the extra cognitive load that comes with analytic scoring technique. Concentrating on many aspects at the same time may affect raters' focus and may divert them from the overall effect of the assessed work (Hughes, 2003; Luoma, 2004). In that case, providing several aspects to assess might inhibit the teachers' concentration who is at the same time listening to the students' performance. O'Sullivan (2008) argues that it would be unfair for the teacher who also acts as the interlocutor and concentrates on many aspects of scoring criteria. This condition somehow might affect the teachers' cognitive capacity and lead to inaccurate evaluation.

Based on those demerits of analytic scoring technique, the researcher purposed peer assessment to avoid the marking load face by the teacher. In this study, the assessment was conducted by the students.

Peer Assessment

Topping (2009) asserted that peer assessment is the assessment process that involves students to judge the other students in equal level of status by considering the value, level and the quality of the product. Product here includes written, oral presentations, portfolios, test performance, or other skilled behaviours. Equally, Spiller (2012) also stated that despite of the type of peer assessment, substantially in this activity, the students provide feedback by recognising the strengths and weaknesses of their peer work. According to Falchikov (2003), "Peer assessment requires students to provide either feedback or grades (or both) to their peers on a product or a performance, based on the criteria of excellence for that product or event which students may have been involved in determining". However, Spiller (2012) claimed that the assigning of grade from the peer has been recognizably difficult to conduct. Falchikov and Goldfinch (2000) state that during the assessment, the students utilize some criteria and standard to facilitate them when reflecting to their peer work.

The Way to Employ Peer Assessment

Spiller (2012) and Falchikov (2003) explored several aspects in employing peer assessment in classroom. Firstly, according to Falchikov (2003), peer assessment will be better with practice. Practice will help the students to be more competent and gain self-confidence. The activity that could facilitate students in conducting peer assessment are exchange and discussion of lecture notes (Spiller, 2012). Secondly, the criteria what to measure or to judge by the students must be stated explicitly and should be understood and can be negotiated with the students if circumstances are appropriate. Falchikov (2003) exemplifies some methods of measurement that can be used such as checklists, rating scales, model answers etc. Thirdly, the implementation of feedback can be summative or formative. Fourthly, the teachers can establish an atmosphere of trust in the classroom circumstance.

Then, the teachers should assure that the learning circumstance should be included with the combination of peer learning and collaboration in a range of ways.

In addition, based on the aspects considered above, Pang (2009) and Laur (2014) summarized the aspect of peer assessment in their studies into three phases. First of all, pre-implementation phase consists of preparation such as preparing the teaching and learning media that could facilitate the activity. The next is training the students to get used to with assessment circumstances. The last is deciding the criteria and measurement to be used during assessment. In the implementation phase, the teacher can demonstrate the implementation of peer assessment, monitor the process and adjust when necessary, and monitor the quality of peer feedback. In post implementation phase, the teacher and students discuss the outcome and reflection or identification of the challenge in implementing peer assessment for better performance and preparation in the future.

Above all, the most important point of implementing peer assessment that can significantly impact the outcome of peer assessment is the type of measurement and its criteria of what needs to assess. In speaking assessment, the technique of measuring students' performance can either use holistic or analytic scoring technique (Heaton, 1989; Brown, 2004).

RESEARCH METHOD

Research Design

This research was intended to know whether or not the use of analytic scoring technique for speaking peer assessment could contribute significant gains on enhancing the students' speaking skill. Therefore, the design applied in this study was experimental research design. Sukardi (2003) claims that experiment research design can be the most productive research design because it can answer the hypothesis which is mainly related to cause and effect.

Due to the incapability of determining every subject of the population to group randomly in the university, the researcher used the available classes as the intact groups. It is in line with Creswell (2012), that in education setting, many experimental occur in a situation in which the researchers use intact group mostly

because of the availability of the participants or because the setting prohibits forming artificial groups. For that reason, this research employed quasi-experimental design with non-equivalent (Pre-Test and Post-Test) control group design. White and Sabarwal (2014) point out that quasi-experimental design includes the formation of two group comparison when it is unlikely or not possible to randomize the participants or groups of the treatment and the control groups.

Research Variable

There were two variables examined in this study namely independent and dependent variable. Fraenkel, Wallen & Hyun (2012) define independent variable as those that the researcher intends to investigate in order to judge their possible effect on one or more variable. Dependent variable, on the other hand, is the variable that independent variable is presumed to affect. Therefore, the present study was aimed at seeking to discern the possible effect between the independent variables, namely (scoring technique for peer assessment), and the dependent variable that is (the students' speaking skill).

Setting and Population Research

1. Setting

This study conducted at University of Merdeka Malang specifically at English Diploma III. The university is located in Jl. Terusan Dieng No. 62-64, Klojen, Pisang Candi, Sukun, East Java, 65146. The place was chosen because in this university provides English Diploma 3 where the students are taking double degree program. In another words, beside concentrating on English major, the students also concentrate on another major existing in this university. Moreover, referring to the result of the preliminary research conducted by the researcher on 15 January 2018 showed that the students in speaking classes struggled to express their thoughts orally.

2. Population research

The target population of this research was all of the students who took the speaking subject in the second semester in the academic year of 2017/2018. There were 43 of students who took the speaking class. As mentioned in the background of study, the accessible population was all of the classes in the second semester.

Because there were only two classes available, population research was used. The population of this research was two speaking classes from second semester of English Diploma 3 of University of Merdeka Malang in academic year 2017/2018. The Speaking A class consisted of 25 students was chosen randomly as control group and the Speaking B class consisted 18 students as experimental group. The way of selecting group was through flipping coin.

Research Instruments

The instruments used in this study were speaking pre-test and post-test, audio recorder and a scoring scale in order to facilitate the students in doing peer assessment. The speaking tests were used as the main sources to determine the homogeneity of the students and to measure the effectiveness of the treatment.

Speaking Tests

There were two speaking tests administrated in this study. Firstly, the speaking pre-test was used to know the homogeneity of the two groups. The speaking test category in this phase used extensive speaking test. Brown (2004) states that extensive speaking test is often used in the assessment in which the speakers produce oral language with limited intervention from the interlocutor such as: speech, oral presentation, picture cued story telling and etc. An oral presentation was selected because this type of activity is available in the students' learning modul for speaking II and had been discussed in the previous meeting. Therefore, the test was considered familiar to the students. In this test, the researcher provided three different topics for oral presentation theme where the students can choose based on their own preference. The topics are 1) promoting a new brand electronic, 2) selling a car and, 3) promoting a hotel. The time allotment for each student was three minute presentation. The detail instruction for the speaking pre-test can be seen in Appendix 1a.

The second was the speaking post-test. The post-test in experimental study was used to measure the students' speaking competence by comparing the result of the post-test of the control group and that of the experimental group. According to Nation and Macalister (2010) the post-test is the achievement test that will be used to know how much the learners have actually learned from in a course and the

effectiveness of the course. In other words, it is used to test the achievement of the objective of the study within a certain period of time (Brown, 2004). The type of speaking post-test and the instruction were similar to those used in the speaking pre-test. The topic and the time allocation were the only things that differentiate between the pre and post-test. There were three topics proposed in post-test which are 1) promoting a homemade product, 2) selling a house and, 3) promoting a tourism place. The detail instruction for the speaking post-test can be seen in Appendix 1b.

Scoring Scale

To assess students' speaking skill, holistic and analytic scoring scales were employed in the study. The researcher adapted the scoring scales that have been developed and validated in Testing Office of EU SFL. Testing office of EU SFL is the testing office that is responsible for test administrations and evaluations of the scores. After designing and utilizing the holistic scoring scale for almost ten years, the Testing Office of EU SFL designed the analytic scoring scale based upon the relevant literature and the needs of the EU SFL. These two types of scoring scales also have been used by Önem (2015) for assessing speaking in Turkish EFL context. Hence, since these scoring scales have been used in EFL context, the scoring scales can be considered as reliable and valid for the present study where it was conducted in Indonesia where English is taught as a foreign language.

However, although the scoring scales had been validated, the implementation of both holistic and analytic scoring scales were adjusted based on the need in the research field or learning objectives of the students. The scoring scales can be seen in Appendix 2a & 2b.

Rating Process

Since the speaking pre-test and post-test are subjective tests, the scoring process can only be done subjectively. Hence, the score depends on the scorer's impression toward the students' speaking performance which are categorized into various aspects such as pronunciation, grammar, vocabulary, fluency and comprehension. Therefore, to avoid the subjectivity, the rating process will be done by two raters. The selected raters are experienced English lecturers who have taught

the Speaking subject for five years. Gatbonton (1999), Martin & Mayall, (2006) Richards & Tang (1998), Tsui (2003) state that experienced teachers, those who have approximately five years or more of classroom experience in teaching English. The reason of choosing two experience raters was to minimize the subjectivity during the scorer since they have familiar with the type of analytic scoring scale used in this study.

The rating process was conducted by measuring students' pre-test and post-test performance that have been recorded by researcher. The maximum acceptable differences between rater one and rater two was 2 point so that there was not any significant difference of the two scorers. When there was more than two-point difference between rater one and two, the researcher asked for clarification and discuss the difference then minimize it. The last, the result from the raters were accumulated then divided into two to get the final score.

Test Validity

Test validity refers to measuring what is supposed to measure (Heaton, 1989, Nation and Macalister, 2010). In other words, a test is considered valid when it is representing what is to evaluate. For example, in speaking assessment, a monologue test has high validity because the students completely produce the oral activity. Simply Therefore, Brown (2004, p. 22) argues that validity is the most crucial element in assessment.

Heaton (1989) breaks down validity into four categories. The first is face validity; it can be described when the test items seems right to other evaluators, teachers, moderators and test takers. Nation and Macalister (2010) illustrate that if the test subject is reading, the test item also supposed to be reading test. If the test is about speaking test, the test should be oral performance. Bachman and Palmer (1996) state that, even in a glance, we can immediately see if the test is appropriate to target subject of evaluation. The second is content validity. It is an indication that determines a certain interpretation of test to be justified. According to Nation and Macalister (2010) content validity is a little similar with face validity, except that the identification in content validity is not just by looking at the test "face" but by analysing the content of the test. Wang, Chen, Gong and Tiura (2009) state a test

has content validity if its content represents the sample of the subject skills and precise structures of skill component. The third is construct validity. Heaton (1989) points out that the test has construct validity when it can fulfil the requirement of a particular feature that is in accordance with the theory of language behaviour and learning. For example, if the test purpose requires on systematic language of structural approach, then a test which emphasizes communicative aspect has low construct validity. Similarly, if the communicative approach to language teaching has been implemented throughout a course, a test with multiple choice type will have lack of construct validity. The last is empirical validity. This validity is obtained as a result of comparing the results of the test with the result of some criterion measure. The test situations are always an important factor in determining the overall validity of any test (Heaton, 1989).

In related to this study, construct validity and content validity were used. In order to prove that the result of the test fulfilled the construct validity, the task was in the form of direct speaking performance test which is oral presentation. The direct speaking performance was selected by considering two aspects, namely: 1) direct speaking test requires students to express their own words orally so that the test measures what it is intended to measure and 2) direct speaking test is easier and quicker to prepare. In addition, for content validity, test can be called to have validity if it contains a representative sample of skills, structure, and so on that it is meant to be concerned (Hughes, 2003). In related to this study, the test was used to measure the students' ability in expressing ideas orally in the form of oral presentation. Thus, the speaking components that were included in the scoring scale used to measure the students' speaking performance consists of pronunciation, grammar, vocabulary, fluency, and comprehension. Therefore, the test can be claimed to have the content validity evidence.

The subjects of this study performed oral presentation with a given topic in three minutes. The speech was in direct speaking performance not by recording because the test situation could not be controlled. To make sure that the speaking test was appropriate, the researcher asked a speaking lecturer who was the teaching Speaking II lecturer in University of Merdeka Malang and the expert from EFL

lecturer of English Department who acted as the researcher's advisers to review the speaking test.

Reliability

Heaton (1989) points out that reliability is a crucial element of a good test or in scale for productive skill. Reliability in testing is the consistency of score on a test (Bachman and Palmer, 1996, Heaton, 1989, Nation and Macalister, 2010). In other words, the consistency of test score refers to the nearly similar result despite the test held in different occasions. For example, in speaking performance, one student who performs oral assessment evaluated by teacher A, will yield nearly the same result if he or she is evaluated by another different rater using similar scoring scale. In that case, it shows that the scoring scale has fulfilled the reliability. Although, there are some possible aspects which might influence the identical result such as fatigue, student error in responding, or even the examination place (Wells & Wollack 2003).

According to Brown (2004) human error, bias and rater subjectivity might happen in the evaluation process. Therefore, reliability is classified into inter-rater reliability that happens when two or more raters get different scores in the same test. For example, in the statistic, another way of measuring the inter-rater reliability among the raters are by deciding a certain category then calculate the agreement. Then, if the raters agree 8 out of 10 times, meaning that the test has 80 % of inter rater reliability (Cherry, 2009). Inequality result commonly occurs due to several factors such as: lack of attention to scoring criteria, inexperience, or preconceived biases (Brown, 2004).

The reliability of this study was seen by the same score given to the students' speaking performance when it was rated by two or more raters. By using inter rater reliability, the consistency of the students' scores will be known. In this study, the researcher chose to apply inter rater reliability since the result can show the consistency of the students' scores under the condition that the other corrector has the same qualification and competency in the field being tested and the scoring was based on the same guidelines. The maximum acceptable difference between rater

one and two is 1 point so that there was not any significant difference from the two raters' scores.

Research Procedure

The procedure of this experimental research involved (1) research procedure for experimental group and (2) research procedure for control group.

Procedure for the Experimental Group

The activities of the teaching and learning process of speaking in the experimental group conducted in eight meetings. These eight meetings conducted in the 9th up to 16th meetings based on the class schedule. The 10th meeting was used for pre-test and introduction to the idea of the analytic scoring technique with the example of simulation how to employ the technique in speaking peer assessment. To optimize the time during experiment, the general description related to analytic and holistic scoring techniques had been introduced by the teacher along with the example of scoring scale in early meeting of the speaking II subject.

After students got accustomed to the use of analytic scoring technique for speaking peer assessment, in the 11th up to 15th meeting, the teacher started the treatment. In this group the researcher employed peer assessment using analytic scoring technique. The students assessed their peers by focusing on several criteria provided in analytic scoring scale that had been provided by the researcher. The analytic scoring technique were distributed before the students performing in front of the class. After one student performed, another students tried to give feedback based on the criteria stated in analytic scoring scale. The more specific detailed related to the activity was elaborated in the lesson plans (appendix 4). The last meeting, 16th meeting is used for the post-test.

Procedure for the Control Group

In the control group, the same material, teaching method, media as those used in the experimental group will be employed. This class also will have eight meetings the same as the experimental class that are based on the class schedule. The 9th meeting will be used as for the pre-test and general introduction to the holistic scoring technique. In the 10th up to the 15th meetings, the same learning activities of the control group conducted as those of the experimental group by

distinction only on the assessment scoring technique where the control group used holistic scoring technique for peer assessment. The students assessed their peers based on overall impression of the performance. In this group the researcher provided holistic scoring scale to facilitate the students during assessment. The holistic scoring technique were distributed before the students performing in front of the class. After one student performed, another students tried to give feedback based on the criteria stated in holistic scoring scale. The more specific detailed related to the activity was elaborated in the lesson plans (Appendixes).

Data Analysis

Dealing with the students' speaking performance, the data analysis was intended to investigate the difference of the students' speaking performance before and after the treatment. Since this study used quasi experimental, non-randomized control group, pre-test-post-test design, the students' speaking achievement prior to the treatment should be involved in the analysis.

Before testing the hypotheses, the researcher tested the normality and the homogeneity of the data collected. According to Larson - Hall (2010), the normality of data should be check before conducting statistical test. In addition, homogeneity testing is a test to measure wheather the data is homoegeneous (Sudjana, 2005). Larson - Hall (2010) suggests the way of testing homogeneity of variance is to use Levene's test. Therefore, the two types of testing above will be done by using statistical program namely IBM SPSS 21.0. The last, hypothesis testing was conducted to investigate wheather or not the analytic scoring technique through speaking peer assessment could enhance the students' speaking skill. Since there were two groups that experience different treatments, the mean of both the two groups were compared using independent sample t-test (Larson - Hall, 2010).

FINDINGS AND DISCUSSION

This chapter conveys the research findings, the verification of the research hypothesis and discussion related to the finding.

Research Findings

As stated in Chapter I, this study was intended to know whether or not the use of analytic scoring technique for speaking peer assessment could contribute significant gains on enhancing the students' speaking skill compared to those assessed by holistic scoring technique.

The Result of the Post-Test

The post-test was administrated to get the data of students' speaking score after employing the different treatment in the experimental and the control group. As elaborated in the preceding chapter, the analytic scoring technique for speaking peer assessment was employed in the experimental group. In the control group, however, the holistic scoring technique for speaking peer assessment was employed. The treatment and the test were conducted based on the arranged schedule. Therefore, the result of the post-test from both of the groups were used as the main data.

Based on the result of the post-test in the experimental and control groups, it was known that the mean score for the experimental was 76.50 and 73.60 for the control group. The summary of difference between the experimental group and the control group is illustrated in Table 4.4.

Table 4.4 The summary difference between the experimental group and the control group

Report					
posttest_result	N	Minimum	Maximum	Mean	Std. Deviation
group class					
Experiment	18	73.00	82.00	76.5000	2.33263
Control	25	70.00	80.00	73.6000	3.52373
Total	43	70.00	82.00	74.8140	3.37537

Based on the table above, the students' speaking post-test scores in the experimental group ranged from 73 to 82 and the standard deviation was 2.33 respectively. Meanwhile, the students' speaking post-test scores in the control group ranged from 70 to 80 and the standard deviation was 3.52 respectively.

Prior to analysing the data for hypothesis testing, homogeneity and normality test was carried out as the fulfilment of the statistical assumptions. The

data used to fulfil the statistical assumption were the data obtained from the result of speaking post-test performance. All the testing was conducted by using SPSS 21.0.

1. Normality of the Data

In order to see the normality of the data, the Kolmogorov Smirnov test was used. The data distributes normally if the observed significance is higher than 5% level of significance, or .05. The result of the test of the normality for speaking post-test score of the experimental and the control group can be seen in table 4.5.

Table 4.5 Test of Normality

One-Sample Kolmogorov-Smirnov Test		Post-test result
N		43
Normal Parameters ^{a,b}	Mean	74.8140
	Std. Deviation	3.37537
Most Extreme Differences	Absolute	.129
	Positive	.129
	Negative	-.101
Kolmogorov-Smirnov Z		.847
Asymp. Sig. (2-tailed)		.470

Based on the data above, it can be seen that the Asymp. Sig (2-tailed) from experimental and control groups showed .470, which was higher than .05. Therefore, the data had fulfilled the normality assumption.

2. Homogeneity of the Data

To find out the information about the homogeneity of the data, the homogeneity test was conducted by using Lavene test in SPSS 21.0. The significant level of the test was .05. In this case, if the result of the analysis is more than .05 then the data are homogeneous, or there is no significance difference between the two groups. However, if the result of the analysis is less than .05, then the data are not homogeneous or there is any significant difference between the experimental group and the control group. The result of the homogeneity test of speaking post-test score from experimental and control group can be seen in Table 4.6.

Table 4.6 Test of Homogeneity using Levene statistic

Test of Homogeneity of Variances			
Levene Statistic	Post-test result		
	df1	df2	Sig.
3.034	1	41	.089

The level of significance of the homogeneity test of the experimental and control groups was .089. It showed that the level of significance of the homogeneity test was higher than that of .05. It indicated that the variances of the data were homogeneous.

All things considered, parametric statistical analysis was used to test hypothesis because all the assumption had been fulfilled. The data were distributed normally and homogeneous. The parametric testing used in this study was independent sample t-test.

3. Hypothesis testing

Testing hypothesis is the main purpose in conducting this research. It was conducted to see whether or not the use of the analytic scoring technique for speaking peer assessment contributed significant gains on enhancing the students' speaking skill. In order to see the first hypothesis, parametric analysis was used independent sample t-test. The t-test was used in order to obtain the differences between the experimental and the control groups since normality and homogeneity of the data had already been fulfilled. The result of the independent sample t-test can be seen in table 4.7.

Table 4.7 The result of post-test using Independent sample t-test

		Independent Samples Test							
		Levene's Test for Equality of Variances				t-test for Equality of Means			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference
									Lower Upper
posttest_result	Equal variances assumed	3.034	.089	3.040	41	.004	2.90000	.95400	.97337 4.82663

Equal
variances
not assumed

From table 4.7, it can be seen that the t value on equal variance assumed was 3.040. Then, the distribution of t table on $\alpha = 5\%$: $2 = 2.5\%$ (2 -Tailed) with degree of freedom (df) $n - 2$ ($43 - 2 = 41$) was 2.01954 (see the appendix 7). Based on the aforementioned table above, t value (3.040) was higher than t table (2.01940). Thus, it showed that the alternative hypothesis was accepted. In other words, Assessing students by using analytic scoring technique for speaking peer assessment contributes significant gains on enhancing the students' speaking skill compared to those assessed by holistic scoring technique.

Moreover, based on the result of the independent sample t -test score above, it was also observed that the obtained probability from the two groups was .004 at the .05 level of significance. Since the probability was less than the .05, there was no meaningful evidence to accept the Null hypothesis. Thus, it can be concluded that assessing students by using analytic scoring technique for speaking peer assessment contributed significant gains on enhancing the students' speaking skill compared to those that assessed by holistic scoring technique.

Discussion

Based on the statistical analysis of the data, the formulated research problem on whether or not the use of analytic scoring technique for speaking peer assessment can contribute significant gains on enhancing the students' speaking skill compared to those assessed by holistic scoring technique was answered. The result revealed that there was any significant difference between the two groups. This finding leads us to conclude that analytic scoring technique clearly contributes significant gains in enhancing students' speaking skill.

This research finding was in line with some other conducted by Moskal (2000), Park (2004) Downing and Haladyna (2006), Tuan (2012) and Hosseini and Mowlaie (2016) who studied about analytic scoring technique. In their study, the

assessment using analytic technique was conducted by the teacher. However, in this study, the assessment and the feedback of the students' performance were conducted by the students as the peer assessment. Demore & Wyoming (2017) also found that analytic scoring technique for students' self-assessment was effective. In their study, self-assessment was administered to facilitate the teacher to have quick and efficient assessment. In that context, the assessment conducted by peer assessment shared the same benefit where the students are not only getting involved with the learning process but also took the benefit of the existing scoring rubric where they can reflect their own performance. Those activities are highly related with contemporary learning approach that emphasizes on the students' engagement on their own learning to promote their responsibility, metacognitive skills and a dialogical, and collaborative model of teaching and learning (Spiller, 2012). Hence, it can be concluded that assessing students' speaking skill using analytic scoring technique can be employed by the teacher, peer assessment or for self-assessment.

Nevertheless, there were also some studies which showed different results, that is analytic scoring outperformed holistic scoring technique. Ounis (2017) stated that regardless of the differences, it is noted that holistic scales are more useful, reliable and consistent as far as speaking judgment is concerned. Metruk (2018) showed there was statistically significant difference between holistic and analytic assessment as a p-value is calculated at 0.001 ($p < .05$). However, this study suggested to employ both techniques of scoring since the limitation of this study only involved one rater during scoring students' performance. Therefore, the subjectivity during the scoring was highly to be avoided.

To sum up, based on reporting to the result of this research, there are several considerations why the alternative hypothesis was accepted and why null hypothesis was rejected. Firstly, the analytic scoring technique significantly enhanced students' speaking performance because the students were guided to pay attention to the detail error of their peer performance. For instance, the students were able to detect the errors from the grammar and pronunciation so that they could discuss them and gave the feedback after the performance. Secondly, the analytic scoring technique was easier to be applied in formative assessment as it provided explicit

criteria that needed to be assessed. In this study, the explicit criteria was presented in the analytic scoring rubric during the assessment, the type of which was formative assessment. According to Satheesh, Brockmann, Cynthia and Amyot (2015), by providing explicit criteria in rubric may help the students who are not aware of the specific aspects for valid assessment.

On the contrary, assessing by using holistic scoring technique inhibited students to have comprehensive evaluation. Hence, the feedback given by peer assessment was limited. Dickins, 2004; Fulcher, 1996b; Kelly & Turner, 2007; Upshur & Turner, 1995 as cited in Jamieson and Poonpon (2013) argue that the holistic scoring is too broad to apprehend speaking improvement of students and hardly facilitate the teacher with satisfactory guidance for formative classroom assessment. The last, since the feedback and evaluation using the holistic scoring technique was limited, the support from the peers were also limited. Nation & Newton (2009) argue that the performance conditions have an impact towards the learners' output.

CONCLUSION AND SUGGESTION

Conclusion

On the basis of the research finding, it can be concluded that there was significant enhancement in the students speaking result after employing speaking peer assessment by using analytic scoring technique. The result of post-test between the experimental group assessed by peer assessment by using the analytic scoring technique outperformed the students' post-test result in control group assessed by peer assessment using holistic scoring technique.

Based on the result of independent sample t-test score, it was observed that t value (3.040) was higher than t table (2.01940). Thus, it showed that the alternative hypothesis was accepted. Seeing from sig.(2 tails) the significance level of the two groups was .004 less than .05. Hence, it can be concluded that the null hypothesis was rejected, and the alternative hypothesis was accepted. In other words, the students' speaking skill of post test result in the experimental group significantly

improved compared with the students' speaking result in the control group assessed by peer assessment using holistic scoring technique.

Suggestions

Reflecting on the conclusion above, this study could propose several suggestions. The first, since the study focused on speaking assessment, the main suggestion was mainly directed to the teacher who are responsible for assessing students' speaking performance in teaching and learning process. In speaking subject, teachers can utilize the analytic scoring technique since it provides detail and explicit criteria of language components that can facilitate the teachers to have valid assessment. Nevertheless, to avoid the marking load, the teachers also can involve the students by providing them with proper analytic scoring scale to facilitate them during assessment.

The second, the result of the study indicated that the analytic scoring technique for peer assessment using analytic scoring technique can improve the students' speaking skill. Hence, students can participate in every aspect of the teaching and learning process including assessment to make them more responsible for their own learning. Moreover, by participating in speaking assessment, the students can be aware about what aspects or criteria are required to have a good speaking skill.

The last, the final suggestion is addressed to the other or future researchers who keen on investigating speaking scoring techniques. There are still several aspects that can be analyzed. For example, the researchers can analyze the challenge of utilizing analytic scoring technique in peer assessment or investigating students' and teachers' attitude towards scoring technique in speaking peer assessment.

REFERENCES

- Alderson, J., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press
- Bacha, N. (2001). Writing evaluation: What can analytic versus holistic essay scoring tell us? *Elsevier Science Ltd. System*, 29 (3), 371–383.

- Bachman, L. & Palmer, A.S.(1996) *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford University Press
- Biggs,J. & Tangs, C. (2009) *Teaching for Quality Learning at University (3rd ed)*. Maidenhead: Open University press
- Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practice*. NY: Pearson Education,Inc.
- Cherry, K. (2009). *What is Realibility?* Retreived from <http://psychology.about.com/od/researchmethods/f/reliabilitydef.htm>
- Creswell, J. W. (2012). *Qualitative inquiry & research design: Choosing among five approaches* (4th ed.). Thousand Oaks, CA: Sage.
- Demore, W. & Wyoming, L. (2017) "Know Thyself: Using Student Self-Assessment to Increase Student Learning Outcomes" (2017). *SMTC Plan B Papers*. 63.
- Downing, S.M. & Haladyna, T.M. (2006). *Handbook of Test Development*. NJ: Lawrence Erlbaum Associates, Inc.
- Fairbairn, J. & Dunlea, J. (2017). *Speaking and Writing Rating Scales Revision*, London, England: British Council
- Falchikov, N., & Goldfinch, J. (2000). *Student peer assessment in higher education: A meta- analysis comparing peer and teacher marks*. Review of Educational Research, 70(3),
- Falchikov, N. (2003). *Involving students in assessment*. *Psychology Learning, and Teaching*, 3(2),
- Febrianti, E. & Rosana, E. (2011) Teaching Speaking of English as A Foreign Language: Problems and Solutions. *Jurnal Bahasa dan Sastra* Vol 1, No 2 (2011)
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to desIgn and evaluate research In educatIon* (8th ed.). New York: Mc Graw Hill.
- Fulcher, G. (2003). *Testing second language speaking*. NY, USA: Taylor & Francis
- Fulcher, G. (2010). *Practical Language Testing*. London: Hodder Education.
- Fulcher, G. & Davidson, F. (2007). *Language testing and assessment*. London & New York: Routledge

- Gatbonton, E. (1999). Investigating experienced ESL teachers' pedagogical knowledge. *Modern Language Journal*, 83(1), 35-50.
- Heaton, J.B (1989). *Writing English Language Testing*. NY: Longman Inc.
- Hosseini, M. & Mowlaie, B (2016) Effect of Holistic vs. Analytic Assessment on Improving Iranian Intermediate EFL Learners' Writing Skill. *Journal of Language and Translation* Volume 6, Number 1(11), (pp.31-41), 2016
- Hughes, A. 2003: *Testing for language teachers*. 2nd edition. Cambridge: Cambridge University Press. 251 pp. ISBN: 0 521 48495 2.
- Huot, B.A. (1990). *Reliability, validity and holistic scoring: what we know and what we need to know*. College Composition and Communication 41,
- Jamieson, J., & Poonpon, K. (2013). Developing Analytic Rating Guides for Toefl Ibt'S Integrated Speaking Tasks. *ETS Research Report Series*, 2013(1), i-93. <https://doi.org/10.1002/j.2333-8504.2013.tb02320.x>
- Laur, K. (2014). Using Peer Assessment In Basic School Efl Classes. *University Of Tartu Department Of English Language And Literature*.
- Larson – Hall, J. (2010). *A Guide to Doing Statistic in Second Language Research Using SPSS*. NY: University of North Texas.
- Lumley, T. (2005). *Assessing second language writing: The raters' perspective*. Frankfurt: Peter Lang.
- Martin, N. & Mayall, H. (2006). *Classroom management training, teaching experience and gender: Do these variables impact teachers' attitudes and beliefs toward classroom management style?* Paper presented at the Annual Conference of the Southwest Educational Research Association, Austin, TX.
- McNamara, T. (2000). *Language Testing*. Oxford: Oxford University Press
- Metruk, R (2018). Comparing Holistic and Analytic Ways Of Scoring in The Assessment Of Speaking Skills. *The Journal Of Teaching English For Specific And Academic Purposes*. Vol. 6, No 1, 2018, Pp. 179-189
- Moskal, B.M. (2000). "Scoring rubrics: What, when and how?" *Practical Assessment, Research & Evaluation*, 7 (3).
- Mukminatin, N. (2000) The advantages of analytic scoring procedure in speaking assessment. *TEFLIN Journal*. Vol.11 No.1 2000

- Nation, I. & Macalister, J. (2010). *Language Curriculum Design*. Routledge, New York, Taylor & Francis e-Library
- Nation, I. & Newton, J. (2009). *Teaching ESL/EFL Listening and Speaking. ESL & Applied Linguistics Professional Series*. Oxfordshire: Routledge Taylor & Francis Group.
- Önem, E. E (2015) Instructors' Attitudes towards Assessing Speaking Holistically and Analytically *Teaching English as a Foreign Language İhsan Doğramacı Bilkent University Ankara*
- O'Sullivan, B. (2008). *Modelling performance in tests of spoken language*. Bern, Switzerland: Peter Lang.
- O'Sullivan, B. (2015b). *Linking the Aptis reporting scales to the CEFR*. TF/2015/003. London: BritishCouncil.
- Ounis, M. (2017). A Comparison between Holistic and Analytic Assessment of Speaking. *Journal of Language Teaching and Research*, 8(4), 679–690.
- Pang, J.C. (2009). Peer assessment of oral presentation in an EFL context. Submitted to the faculty of the University Graduate School in partial fulfilment of the requirements for the degree Doctor of Philosophy in the Department of Literacy, Culture, and Language Education Indiana University. Available at <http://gradworks.umi.com/3380148.pdf>,
- Park, T. (2004). *An Overview of Portfolio-based Writing Assessment. Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, Vol. 4, No. 2 The Forum
- Richards, J. & Tang, A. (1998). *Exploring pedagogical reasoning skills*. In J. C. Richards (Ed.), *Beyond training: Perspectives on language teacher education* (pp. 86-102). NY: Cambridge University Press.
- Satheesh, K. Brockmann, L. Cynthia, Y. and Amyot, G (2015). Use of an Analytical Grading Rubric for Self-Assessment: A Pilot Study for a Periodontal Oral Competency Examination in Predoctoral Dental Education. *Journal of Dental Education* December 2015, 79 (12) 1429-1436;
- Solcova, P. (2011). Teaching Speaking Skills. Thesis of Faculty of Arts of MasarykUniversity: Unpublished.
- Sudjana, N. 2005. *Dasar-dasar Proses\Belajar Mengajar*. Bandung. Sinar Baru Algensindo.

- Sukardi, (2003). *Metodologi penelitian pendidikan kompetensi dan praktiknya*. Jakarta: PT. Raja Grafindo Persada
- Spiller, D. (2012). *Assessment Matters: Self-Assessment and Peer Assessment Teaching Development*: Wāhanga Whakapakari Ako
- Topping, K. (2009). Peer Assessment. Theory into Practice. *Taylor & Francis* 48, 20-27. doi:10.1080/0040580802577569
- Torky, S. (2006) The Effectiveness of a Task- Based Instruction program in Developing the English Language Speaking Skills of Secondary Stage Students. *Curricula and Methods of teaching Department*.
- Tsui, A. (2003). *Understanding expertise in teaching: Case studies of ESL teachers*. NY: Cambridge University Press.
- Tuan, L. (2012). Teaching and Assessing Speaking Performance through Analytic Scoring Approach. *Theory and Practice in Language Studies*, 2(4), 673–679.
- Wang, P., Chen, X., Gong, J. and Tiura, A., (2009). Reliability and Validity Personal Social Capital Scale: *An instrument for health and behavioral research*. *Health Education Research*, 24 (2), 306–317.
- Weigle, S. (2002). *Assessing Writing*. Cambridge: Cambridge University Press
- Wells, C. & J. Wollack, (2003). *An Instructor's Guide to Understanding Test Reliability, Testing & Evaluation Services*. University of Wisconsin: Madison
- White, H. & S. Sabarwal (2014). Quasi-experimental Design and Methods, Methodological Briefs: Impact Evaluation 8. *UNICEF Office of Research*

APPENDIX 1: SPEAKING TEST

APPENDIX 1a: Speaking Pre-Test

The test is in the form of individual oral presentation

I. Instruction

Choose one the following topics provided below and present it orally in front of the class for a 3 (three) minutes.

II. The Topics

1. Promoting a New Brand Electronic
2. Selling a Car
3. Promoting a Hotel



APPENDIX 1b: Speaking Post-Test

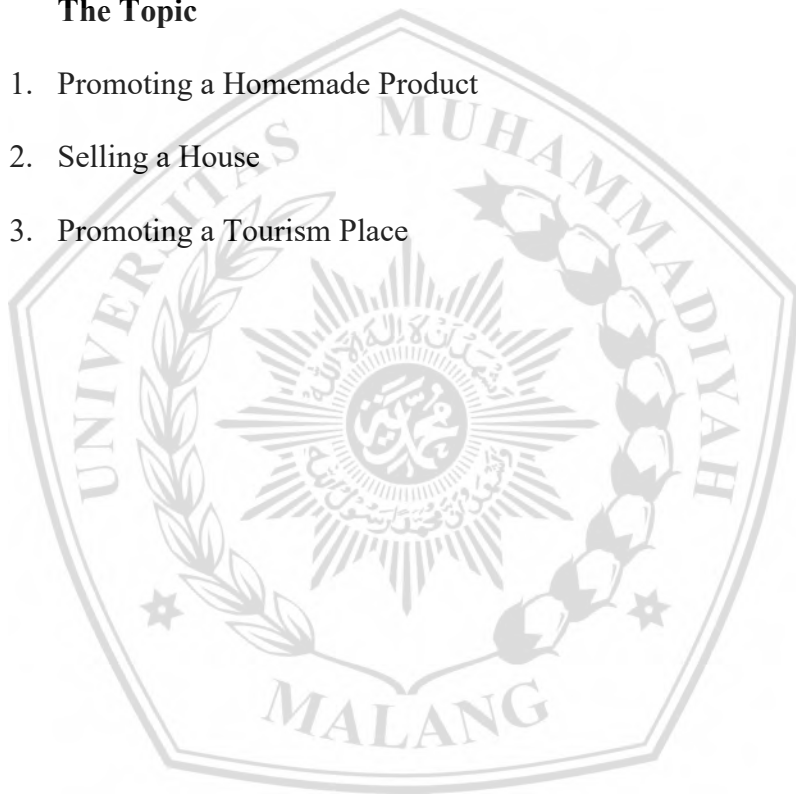
The test is in the form of individual oral presentation

I. Instruction

Choose one the following topics provided below and present it orally in front of the class for a 3 (three) minutes.

II. The Topic

1. Promoting a Homemade Product
2. Selling a House
3. Promoting a Tourism Place



APPENDIX 2 : SCORING SCALE

APPENDX 2a : The Analytic Scoring Scale for Peer Assessment in The Experimental Group and for Scoring Speaking Pre-Test and Post-Test

The analytic scoring scale is adapted from Önem (2015)

	Pronunciation	Grammar	Vocabulary	Fluency	Comprehension
Very poor (0-4)	Major errors in Pronunciation	Major errors in grammar	Very limited knowledge/use of English word	Hesitation and/or pause all the time	Almost no or too little understanding
Poor (5-8)	Frequent errors in pronunciation	Frequent errors in grammar	Frequent errors in word choice	Frequent hesitation and/or pause	Limited understanding and/or needs constant repetition and rephrasing
Average (9-12)	Occasional errors in pronunciation	Occasional errors in constructing structures/grammar	Occasional errors in word choice	Occasional hesitation and/or pause	Understand carefully and/or simplified speech with fewer repetition
Good (13-16)	Minor errors in pronunciation	Minor problems in grammar	Minor errors in word choice	Minor hesitation and/or pause	Understand normal and/or simplified speech with fewer repetition
Very Good (17-20)	Few errors or no error in pronunciation	Few errors or no error in grammar	Effective/ appropriate word choice	No hesitation and /or pause	Understand everything without repetition

Appendix 2b: The holistic scoring scale for speaking peer assessment in the control group

The holistic scoring scale is adapted from Önem (2015)

Holistic oral assessment criteria		
Criteria	Description	Score
Very good	Communicate effectively and accurately with minor mistakes. Fluent and no assistance necessary	86-100
Good	Communicate well with few mistakes. Maintain effectively her/his own part of discussion	71-85
Average	Communicate with some mistakes. Need a little effort on the part of the listener	51-70
Poor	Has problem in communication. Frequently inaccurate, little meaning conveyed	25-50
Very poor	Has serious problem in communication. Limited or no understanding	0-25

APPENDIX 3: LESSON PLAN

Appendix 3a : Lesson Plan of Experimental Group

Lesson Plan

Course	: Speaking II
Topic	: Making conversation part 1 and II
Credit	: 3 credits
Study Program	: English Diploma III
Semester	: 2
Time allotment	: 2 x 60 Minutes
Meeting	: 9-10

I. Basic Competence (s)

Providing students with dialog and monolog based on social interaction in the daily life conversation.

II. Goal

By the end of this course, students are expected to speak English fluently and are able to apply speaking strategies based on social interaction in daily life.

III. Indicators

1. Identifying the common expressions used in making conversation
2. Identifying the common responses used in making conversation
3. Creating short dialogue
4. Performing the dialogue

IV. Instructional Objectives

In the end of the lesson, the students are expected to:

1. Students are able to mention the common expression used in making conversation
2. Students are able to mention the common responses used in making conversation
3. Students are able to create short dialogue

4. Students are able to perform the dialogue

V. Instructional Materials

- **Conversation Starters & Greetings**
- **Typical Responses**

VI. Media and Sources

1. Learning Media

- Laptop +LCD Projector
- Power Point Presentation
- Whiteboard
- Pictures
- Video

2. Learning Source

- Internet Source <http://englishharmony.com/small-talk/#starters>
- Speaking Module
- Youtube.com

VII. Teaching Method and Strategies

1. Method: Communicative language teaching

2. Strategy: Role play

VIII. Teaching Activities Procedure

- Pre- teaching (10 minutes)
 - Greeting
 - Checking the readiness of student to start the lesson
 - Introducing the topic of the day
- Whilst teaching (90 minutes)
 - Lecturer explains the purpose of making conversation in English
 - Lecturer asks to the students about the possible common expression of making conversation in English they have already known.
 - Lecturer explains the common expression and its functions in making conversation
 - Lecturer asks to the students about the possible common responses of making conversation in English they have already known.
 - Lecturer explains the common responses and its function in making conversation

- Lecturer shows the short video about making conversation in English
 - Lecturer divides the students into five groups consist of 4 or 5 students.
 - Lecturer gives the topic to every group
 - Students prepare their material to share orally about 10 minutes
 - Lecturer gives the analytic scoring scale for each member of the groups
 - Each group performs based on the chosen topic in front the class
 - Other groups assess the peer performance using the analytic scoring scale
 - The students share the feedback related to one student's performance
 - Lecturer monitors the quality of the feedback from each group
- Post teaching (20 minutes)
- Lecturer asks the students' understanding related to the material
 - Lecturer asks the challenge of implementing peer assessment
 - Lecturing and students discuss the effectiveness of the peer feedback
 - Lecturer asks about the challenge of using analytic scoring scale

IX. Assessment

The procedure of assessment is given through the student's peer assessment during group presentation. To make sure whether the objectives of the lesson have been already achieved, process of assessment does not result in scores, but reports on students' learning progress.

Appendix 3a: Lesson Plan of Experimental Group

Lesson Plan

Course	: Speaking II
Topic	: Making Excuse
Credit	: 3 credits
Study Program	: English Diploma III
Semester	: 2
Time allotment	: 2 x 60 Minutes

Meeting : 11

I. Basic Competence (s)

Providing students with dialog and monolog based on social interaction in the daily life conversation.

II. Goal

By the end of this course, students are expected to speak English fluently and are able to apply speaking strategies based on social interaction in daily life.

III. Indicators

1. Identifying the common expressions used in making excuses
2. Identifying the common responses used in making excuses
3. Creating short dialog consists of making excuse context
4. Performing the dialog consists of making excuse context

IV. Instructional Objectives

In the end of the lesson, the students are expected to:

1. Students are able to mention the common expression used in making excuses
2. Students are able to mention the common responses used in making excuses
3. Students are able to create short dialogue consists of making excuse context
4. Students are able to perform the dialogue consists of making excuse context

V. Instructional Materials

- Common expression of making excuse
- Conversation example

VI. Media and Sources

3. Learning Media
 - Laptop +LCD Projector
 - Power Point Presentation

- Whiteboard
 - Pictures
 - Video
4. Learning Source
- Internet Source <https://content.cambly.com/2016/08/08/lesson-28-making-excuses/>
 - Speaking Module
 - Youtube.com

VII. Teaching Method and Strategies

3. Method: Communicative language teaching

4. Strategy: Think pair and share

VIII. Teaching Activities Procedure

- Pre- teaching (10 minutes)
 - Greeting
 - Checking the readiness of student to start the lesson
 - Introducing the topic of the day
- Whilst teaching (90 minutes)
 - Lecturer explains the purpose of making excuse
 - Lecturer asks to the students about the possible common expression of making excuse in English they have already known.
 - Lecturer explains the common expression of making excuse and its functions
 - Lecturer shows the short video about making conversation in English
 - Lecturer divides the students into five groups consist of 4 or 5 students.
 - Lecturer gives the topic to every group
 - Students prepare their material to share orally about 10 minutes
 - Lecturer gives the analytic scoring scale for each member of the groups
 - Each group performs based on the chosen topic in front the class
 - Other groups assess the peer performance using the analytic scoring scale
 - The students share the feedback related to one student's performance
 - Lecturer monitors the quality of the feedback from each group

- Post teaching (20 minutes)
 - Lecturer asks the students' understanding related to the material
 - Lecturer asks the challenge of implementing peer assessment
 - Lecturing and students discuss the effectiveness of the peer feedback
 - Lecturer asks about the challenge of using analytic scoring scale

IX. Assessment

The procedure of assessment is given through the student's peer assessment during group presentation. To make sure whether the objectives of the lesson have been already achieved, process of assessment does not result in scores, but reports on students' learning progress.

Appendix 3a: Lesson Plan of Experimental Group

Lesson Plan	
Course	: Speaking II
Topic	: Telephone Language
Credit	: 3 credits
Study Program	: English Diploma III
Semester	: 2
Time allotment	: 2 x 60 Minutes
Meeting	: 12

I. Basic Competence (s)

Providing students with dialog and monolog based on social interaction in the daily life conversation.

II. Goal

By the end of this course, students are expected to speak English fluently and are able to apply speaking strategies based on social interaction in daily life.

III. Indicators

1. Identifying the common expressions used in telephone

2. Identifying the common responses used in telephone
3. Creating short dialogue about telephone
4. Performing the dialogue about telephone

IV. Instructional Objectives

In the end of the lesson, the students are expected to:

5. Students are able to mention the common expression used in telephone
6. Students are able to mention the common responses used in telephone
7. Students are able to create short dialogue about telephone
8. Students are able to perform the dialogue about telephone

V. Instructional Materials

- **Telephone conversation with some of the key phrases:**
- **Introducing Yourself**
- **Asking Who Is on the Telephone**
- **Asking for Someone**
- **When Someone Is Not Available**
- **Taking a Message**

VI. Media and Sources

1. Learning Media
 - Laptop +LCD Projector
 - Power Point Presentation
 - Whiteboard
 - Pictures
 - Video
2. Learning Source
 - Internet Source <https://content.cambly.com/2016/08/08/lesson-28-making-excuses/>
 - Speaking Modul
 - Youtube.com

VII. Teaching Method and Strategies

Method: Communicative language teaching

Strategy: Demonstration

VIII. Teaching Activities Procedure

- Pre- teaching (10 minutes)

- Greeting
- Checking the readiness of student to start the lesson
- Introducing the topic of the day
- Whilst teaching (90 minutes)
 - Lecturer explains the purpose of making excuse
 - Lecturer asks to the students about the possible common expression of making excuse in English they have already known.
 - Lecturer explains the common expression of making excuse and its functions
 - Lecturer shows the short video about making conversation in English
 - Lecturer divides the students into five groups consist of 4 or 5 students.
 - Lecturer gives the topic to every group
 - Students prepare their material to share orally about 10 minutes
 - Lecturer gives the analytic scoring scale for each member of the groups
 - Each group performs based on the chosen topic in front the class
 - Other groups assess the peer performance using the analytic scoring scale
 - The students share the feedback related to one student's performance
 - Lecturer monitors the quality of the feedback from each group
- Post teaching (20 minutes)
 - Lecturer asks the students' understanding related to the material
 - Lecturer asks the challenge of implementing peer assessment
 - Lecturing and students discuss the effectiveness of the peer feedback
 - Lecturer asks about the challenge of using analytic scoring scale

IX. Assessment

The procedure of assessment is given through the student's peer assessment during group presentation. To make sure whether the objectives of the lesson have been already achieved, process of assessment does not result in scores, but reports on students' learning progress.

Appendix 3a : Lesson Plan of Experimental Group

Lesson Plan

Course : Speaking II

Topic	: Making Appointment
Credit	: 3 credits
Study Program	: English Diploma III
Semester	: 2
Time allotment	: 2 x 60 Minutes
Meeting	: 13

I. Basic Competence (s)

Providing students with dialog and monolog based on social interaction in the daily life conversation.

II. Goal

By the end of this course, students are expected to speak English fluently and are able to apply speaking strategies based on social interaction in daily life.

III. Indicators

1. Identifying the common expressions used in making appointment
2. Identifying the common responses used in making appointment
3. Creating short dialogue in making appointment
4. Performing the dialogue in making appointment

IV. Instructional Objectives

In the end of the lesson, the students are expected to:

1. Students are able to mention the common expression used in making appointment
2. Students are able to mention the common responses used in making appointment
3. Students are able to create short dialogue in making appointment
4. Students are able to perform the dialogue in making appointment

5. Instructional Materials

- **Making appointment in conversation**

6. Media and Sources

5. Learning Media

- Laptop +LCD Projector
- Power Point Presentation
- Whiteboard
- Pictures
- Video

6. Learning Source

- Internet Source
<https://www.eslfast.com/robot/topics/employment/employment15.htm>
- Speaking Module
- Youtube.com

7. Teaching Method and Strategies

5. Method: scientific approach

6. Strategy: Role play

8. Teaching Activities Procedure

- Pre- teaching (10 minutes)
 - Greeting
 - Checking the readiness of student to start the lesson
 - Introducing the topic of the day
- Whilst teaching (90 minutes)
 - Lecturer explains the purpose of making appointment in English
 - Lecturer asks to the students about the possible common expression of making appointment in English they have already known.
 - Lecturer explains the common expression and its functions in making appointment
 - Lecturer asks to the students about the possible common responses of making appointment in English they have already known.
 - Lecturer explains the common responses and its function in making appointment
 - Lecturer shows the short video about making appointment in English

- Lecturer divides the students into five groups consist of 4 or 5 students.
 - Lecturer gives the topic consists of making appointment to every group
 - Students prepare their material to share orally about 10 minutes
 - Lecturer gives the analytic scoring scale for each member of the groups
 - Each group performs based on the chosen topic in front the class
 - Other groups assess the peer performance using the analytic scoring scale
 - The students share the feedback related to one student's performance
 - Lecturer monitors the quality of the feedback from each group
- Post teaching (20 minutes)
- Lecturer asks the students' understanding related to the material
 - Lecturer asks the challenge of implementing peer assessment
 - Lecturing and students discuss the effectiveness of the peer feedback
 - Lecturer asks about the challenge of using analytic scoring scale

9. Assessment

The procedure of assessment is given through the students peer assesment during group presentation. To make sure whether the objectives of the lesson have been already achieved, process of assessment does not result in scores, but reports on students' learning progress.

Appendix 3a : Lesson Plan of Experimental Group

Lesson Plan

Course	: Speaking II
Topic	: Offers and Requests
Credit	: 3 credits
Study Program	: English Diploma III
Semester	: 2

Time allotment : 2 x 60 Minutes

Meeting : 14

I. Basic Competence (s)

Providing students with dialog and monolog based on social interaction in the daily life conversation.

II. Goal

By the end of this course, students are expected to speak English fluently and are able to apply speaking strategies based on social interaction in daily life.

III. Indicators

1. Identifying the common expressions how to make offers and requests
2. Identifying the common responses how to make offers and requests
3. Creating short dialogue about how to make offers and requests
4. performing the dialogue about how to make offers and requests

IV. Instructional Objectives

In the end of the lesson, the students are expected to:

1. Students are able to mention the common expression how to make offers and requests
2. Students are able to mention the common responses how to make offers and requests
3. Students are able to create short dialogue about how to make offers and requests
4. Students are able to perform the dialogue about how to make offers and requests

V. Instructional Materials

- **Making requests - asking if you can do something**
- **Offering to do something for another person**

VI. Media and Sources

7. Learning Media

- Laptop +LCD Projector
- Power Point Presentation
- Whiteboard
- Pictures
- Video

8. Learning Source

- Internet Source <http://englishharmony.com/small-talk/#starters>
- Speaking Module
- Youtube.com

VII. Teaching Method and Strategies

7. Method: scientific approach

8. Strategy: Demonstration

VIII. Teaching Activities Procedure

- Pre- teaching (10 minutes)
 - Greeting
 - Checking the readiness of student to start the lesson
 - Introducing the topic of the day
- Whilst teaching (90 minutes)
 - Lecturer explains the purpose of making offers and requests in English
 - Lecturer asks to the students about the possible common expression of making offers and requests in English they have already known.
 - Lecturer explains the common expression and its functions in making offers and requests
 - Lecturer asks to the students about the possible common responses of making offers and requests English they have already known.
 - Lecturer explains the common responses and its function in making offers and requests
 - Lecturer shows the short video about making offers and requests in English
 - Lecturer divides the students into five groups consist of 4 or 5 students.
 - Lecturer gives the topic to every group

- Students prepare their material to share orally about 10 minutes
 - Lecturer gives the analytic scoring scale for each member of the groups
 - Each group performs based on the chosen topic in front the class
 - Other groups assess the peer performance using the analytic scoring scale
 - The students share the feedback related to one student's performance
 - Lecturer monitors the quality of the feedback from each group
- Post teaching (20 minutes)
- Lecturer asks the students' understanding related to the material
 - Lecturer asks the challenge of implementing peer assessment
 - Lecturing and students discuss the effectiveness of the peer feedback
 - Lecturer asks about the challenge of using analytic scoring scale

IX. Assessment

The procedure of assessment is given through the student's peer assessment during group presentation. To make sure whether the objectives of the lesson have been already achieved, process of assessment does not result in scores, but reports on students' learning progress.

Appendix 3a: Lesson Plan of Experimental Group

Lesson Plan

Course	: Speaking II
Topic	: Journal Reading Relay
Credit	: 3 credits
Study Program	: English Diploma III
Semester	: 2
Time allotment	: 2 x 60 Minutes
Meeting	: 15

I. Basic Competence (s)

Providing students with dialog and monolog based on social interaction in the daily life conversation.

II. Goal

By the end of this course, students are expected to speak English fluently and are able to apply speaking strategies based on social interaction in daily life.

III. Indicators

1. Identifying the difficult vocabulary appears in the journal
2. Practicing the pronunciation of difficult vocabulary
3. Summarizing the content of journal orally

IV. Instructional Objectives

In the end of the lesson, the students are expected to:

1. Students are able to know the difficult vocabulary commonly appears in the journal
2. Students are able to know the pronunciation of difficult vocabulary appears in the journal
3. Students are able to summarize the journal content orally used their own words.

V. Instructional Materials



VI. Media and Sources

9. Learning Media

- Laptop +LCD Projector
- Power Point Presentation
- Whiteboard
- Pictures
- Video

10. Learning Source

- Internet Source <http://jurnal.unmer.ac.id/index.php/enjourme>
Speaking Module

VII. Teaching Method and Strategies

9. Method: Communicative language teaching

10. Strategy: Presentation

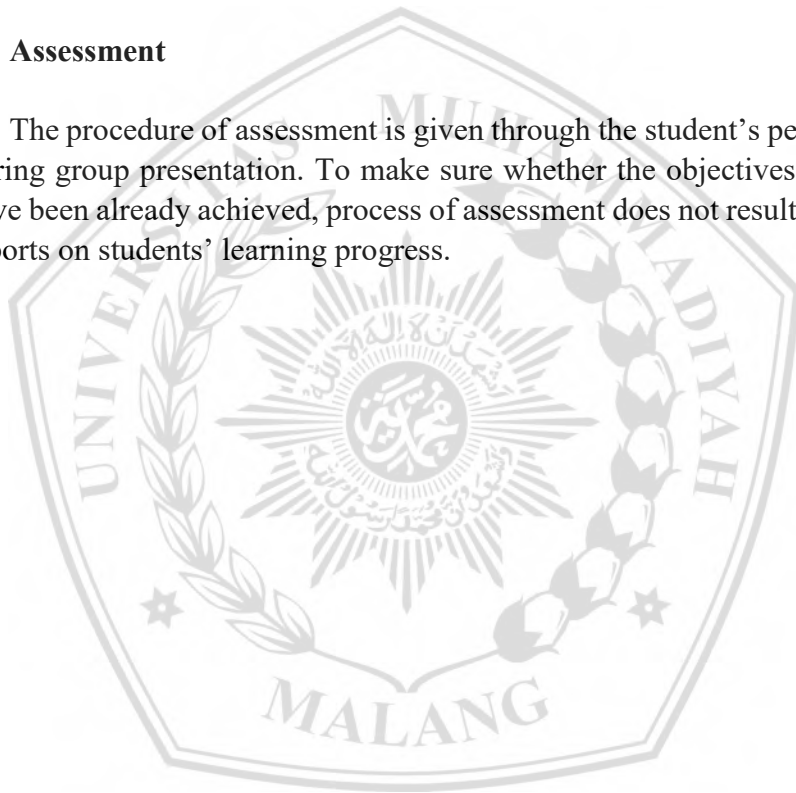
VIII. Teaching Activities Procedure

- Pre- teaching (10 minutes)
 - Greeting
 - Checking the readiness of student to start the lesson
 - Introducing the topic of the day
- Whilst teaching (90 minutes)
 - Lecturer introduce the site of UNMER journal
 - Lecturer asks the students to visit the site using mobile phone or laptop
 - Lecturer explains the student about how to find the journal in site
 - Students download the latest article of journal.
 - Students underline the difficult vocabulary of the journal article
 - Students practice the pronunciation of the difficult words
 - Lecturer asks the students to read the journal article loudly turn by turn
 - Lecturer divides the students into five groups consist of 4 or 5 students.
 - Lecturer gives the topic of the article taken from the journal site
 - Lecturer asks the students to summarize the journal orally
 - Students prepare their material to share orally about 10 minutes
 - Lecturer gives the analytic scoring scale for each member of the groups
 - Each group performs based on the chosen topic in front the class

- Other groups assess the peer performance using the analytic scoring scale
 - The students share the feedback related to one student's performance
 - Lecturer monitors the quality of the feedback from each group
- Post teaching (30 minutes)
- Lecturer asks the students' understanding related to the material
 - Lecturer asks the challenge of implementing peer assessment
 - Lecturing and students discuss the effectiveness of the peer feedback
 - Lecturer asks about the challenge of using analytic scoring scale

IX. Assessment

The procedure of assessment is given through the student's peer assessment during group presentation. To make sure whether the objectives of the lesson have been already achieved, process of assessment does not result in scores, but reports on students' learning progress.



APPENDIX 4 : STUDENTS' ATTENDANCE LIST

APPENDIX 4a : STUDENTS' ATTENDANCE LIST of EXPERIMENTAL GROUP

PRESENTASI KULIAH

Semester Genap 2017 -2018

UNIVERSITAS MERDEKA MALANG

FAKULTAS ILMU SOSIAL DAN ILMU POLITIK

Gedung Kantor Pusat Lt.1 Jl Terusan Raya Dieng 62-64 Malang

Telp.0341-568395 Ext.325. Fax : 0341 564994

website: www.unmer.ac.id, email : pmb@unmer.ac.id

Mata Kuliah : SPEAKING 1 (3)

Kode : 631311

Prodi : Prodi D3 Bahasa Inggris

Ruang : D3BI 301

Kelas : B

SKS : 3

Semester : 2

Kurikulum : 2017

Hari/Jam : Rabu/ 14.30-17.00

Hari/Jam : /-

NO	NIM	Tanggal								
		Nama Mahasiswa	9	10	11	12	13	14	15	16
1	17063000040	CBC	√	√	√	√	√	√	√	√
2	17063000051	DN	√	√	√
3	17063000058	PBS	√	√	√	√	√	√	√	√
4	17063000059	DAL	√	√	√	√	√	√	√	√
5	17063000063	KDA	√	√	√	√	√	√	√	√
6	17063000064	EPF	√	√	√	√	√	i	√	√

7	17063000065	LOM	√	.	√	√	√	i	√	√
8	17063000066	WF		√						
9	17063000007	AAL	√	√	√	√	√	√	√	√
10	17063000030	BW	√	√	√	√	√	√	√	√
11	17063000014	MA	√	√	√	√	√	√	√	√
12	17063000003	LM								
13	17063000034	FP								
14	17063000033	EF	√	√	√	√	√	√	√	√
15	17063000032	ADG	√	√	√	√	√	√	√	√
16	17063000009	DJ	√	√	√	√	√	√	√	√
17	17063000004	RP	√	√	√	√	√	√	√	√
18	17063000062	RP	√	i	√	√	√	√	√	√
19	17063000015	PM	√	√	√	√	√	√	√	√
20	17063000023	CCP	√	√	√	√	√	√	√	√
21	17063000025	HR								
23	17063000027	DB	√	√	√	√	√	√	√	√
24	17063000028	IWW	√	√	√	√	√	√	√	√

KETERANGAN

MAHASISWA HADIR

√

MAHASISWA IJIN (ADA SURAT)

i

MAHASISWA SAKIT

S

MAHASISWA ABSEN

.

APPENDIX 4 : STUDENTS' ATTENDANCE LIST

APPENDIX 4b : STUDENTS' ATTENDANCE LIST of CONTROL GROUP

PRESENTASI KULIAH

Semester Genap 2017 -2018

UNIVERSITAS MERDEKA MALANG

FAKULTAS ILMU SOSIAL DAN ILMU POLITIK

Gedung Kantor Pusat Lt.1 Jl Terusan Raya Dieng 62-64 Malang

Telp.0341-568395 Ext.325. Fax : 0341 564994

website: www.unmer.ac.id, email : pmb@unmer.ac.id

Mata Kuliah : SPEAKING 2

Kode : 631311

Prodi : Prodi D3 Bahasa Inggris

Ruang : D3BI 301

Kelas : A

SKS : 3

Semester : 2

Kurikulum : 2017

Hari/Jam : Senin/ 14.30-17.00

Hari/Jam : /-

NO	NIM	Tanggal								
		Nama Mahasiswa	9	10	11	12	13	14	15	16
1	17063000035	LCF	√	√	√	√	√	√	√	√
2	17063000036	JMS	√	√	√	√	√	√	√	√
3	17063000037	AFI	√	√	√	√	√	√	√	√
4	17063000038	AWS	√	√	√	√	√	√	√	√
5	17063000042	JN	√	√	√	√	√	√	√	√
6	17063000043	AKL	√	√	√	√	√	√	√	√
7	17063000044	ME	√	√	√	√	√	√	√	√
8	17063000046	NI	√	√	√	√	√	√	√	√

9	17063000048	AAC	√	√	√	√	√	√	√	√
10	17063000049	RSH	√	√	√	√	√	√	√	√
11	17063000050	DDT	√	√	√	√	√	√	√	√
12	17063000054	YGD	√	√	√	S	√	.	√	√
13	17063000055	VY	√	√	√	S	√	√	√	√
14	17063000056	RU	√	√	√	√	√	√	√	√
15	17063000057	RA	√	√	√	√	√	√	√	√
16	17063000060	ARA	√	√	√	√	√	√	√	√
17	17063000011	FB	√	√	√	√	√	√	√	√
18	17063000005	AN	√	√	√	√	√	√	√	√
19	17063000009	VA	√	√	√	√	√	√	√	√
20	17063000013	LH	√	√	√	√	√	√	√	√
21	17063000016	REH	√	√	√	.	√	.	√	√
22	17063000020	VAY	√	√	√	√	√	√	√	√
23	17063000006	DSP								
24	17063000012	GG	√	√	√	√	√	√	√	√
25	17063000010	ETD	√	√	√	√	√	√	√	√
26	17063000008	AFAS	√	√	√	√	√	√	√	√
27	17063000031	ADI								

KETERANGAN

MAHASISWA HADIR	√
MAHASISWA IJIN (ADA SURAT)	i
MAHASISWA SAKIT	S
MAHASISWA ABSEN	.

APPENDIX 5 : RESULT OF PRE-TEST

APPENDIX 5A : PRE-TEST OF CONTROL GROUP

NO	NAME	PRE-TEST		AVERAGE
		RATER 1	RATER 2	
1	LCF	70	70	70
2	JMS	68	70	69
3	AFI	78	78	78
4	AWS	72	74	73
5	JN	78	78	78
6	AKL	78	78	78
7	ME	74	72	73
8	NI	72	72	72
9	AAC	69	69	69
10	RSH	69	71	70
11	DDT	68	70	69
12	YGD	68	68	68
13	VY	69	69	69
14	RU	72	72	72
15	RA	73	71	72
16	ARA	71	73	72
17	FB	69	71	70
18	AN	71	69	70
19	VA	72	72	72
20	LH	72	72	72
21	REH	69	71	70

22	VAY	70	70	70
23	GG	72	72	72
24	ETD	78	78	78
25	AFAS	78	78	78

APPENDIX 5 : RESULT OF PRE-TEST

APPENDIX 5a : PRE-TEST OF EXPERIMENTAL GROUP

NO	NAME	PRE-TEST		AVERAGE
		RATER 1	RATER 2	
1	CBC	67	69	68
2	PBS	73	73	73
3	DAL	73	75	74
4	KDA	78	78	78
5	EPF	68	68	68
6	LOM	69	69	69
7	AAL	71	69	70
8	BW	73	73	73
9	MA	73	73	73
10	EF	70	70	70
11	ADG	71	71	71
12	DJ	70	68	69
13	RP	68	68	68
14	RPP	74	74	74
15	PM	72	72	72

16	CCP	72	72	72
17	DB	73	71	72
18	IWW	72	72	72

APPENDIX 5 : RESULT OF POST-TEST

APPENDIX 5B : POST-TEST OF CONTROL GROUP

NO	NAME	POST-TEST		AVERAGE
		RATER 1	RATER 2	
1	LCF	70	70	70
2	JMS	69	71	70
3	AFI	80	80	80
4	AWS	72	74	73
5	JN	78	78	78
6	AKL	79	81	80
7	ME	75	77	76
8	NI	72	72	72
9	AAC	73	75	74
10	RSH	69	71	70
11	DDT	69	71	70
12	YGD	70	70	70
13	VY	75	73	74
14	RU	74	74	74
15	RA	74	76	75
16	ARA	71	73	72

17	FB	71	73	72
18	AN	72	72	72
19	VA	74	74	74
20	LH	72	72	72
21	REH	69	71	70
22	VAY	70	70	70
23	GG	72	72	72
24	ETD	80	80	80
25	AFAS	80	80	80

APPENDIX 6 : RESULT OF POST-TEST

APPENDIX 6B : POST-TEST OF EXPERIMENTAL GROUP

NO	NAME	POST-TEST		AVERAGE
		RATER 1	RATER 2	
1	CBC	75	75	75
2	PBS	79	81	80
3	DAL	75	75	75
4	KDA	82	82	82
5	EPF	73	73	73
6	LOM	76	74	75

7	AAL	75	75	75
8	BW	79	77	78
9	MA	78	78	78
10	EF	76	74	75
11	ADG	77	75	76
12	DJ	75	75	75
13	RP	76	74	75
14	RP	80	80	80
15	PM	75	75	75
16	CCP	79	77	78
17	DB	76	76	76
18	IWW	75	77	76

APPENDIX 6 : STATISTICAL ANALYSIS

Appendix 6 a : Statistical Analysis in Pre-test

Pre-test score

Deskriptive statistic

Report

pretest_result

goup_class	N	Minimum	Maximum	Mean	Std. Deviation
Experimental	18	68,00	78,00	71,4444	2,61719
Control	25	68,00	78,00	72,1600	3,27465
Total	43	68,00	78,00	71,8605	3,00461

NPAR TESTS

/K-S(NORMAL)=pretest_result

/MISSING ANALYSIS.

One-Sample Kolmogorov-Smirnov Test

		pretest_result
N		43
Normal Parameters ^{a,b}	Mean	71,8605
	Std. Deviation	3,00461
Most Extreme Differences	Absolute	,179
	Positive	,179
	Negative	-,119
Kolmogorov-Smirnov Z		1,175
Asymp. Sig. (2-tailed)		,127

a. Test distribution is Normal.

b. Calculated from data.

T-Test

Group Statistics

	goup_class	N	Mean	Std. Deviation	Std. Error Mean
pretest_result	Experimental	18	71,4444	2,61719	,61688
	Control	25	72,1600	3,27465	,65493

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
pretest_result	Equal variances assumed	,495	,486	-,767	41	,448	-,71556	,93338	-2,60055	1,16944
	Equal variances not assumed			-,795	40,487	,431	-,71556	,89971	-2,53325	1,10214

Appendix 6B : Statistical Analysis in Post-test

Post-test score

Descriptive statistic

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
posttest_result * group_class	43	100,0%	0	0,0%	43	100,0%

Report

posttest_result

group_class	N	Minimum	Maximum	Mean	Std. Deviation
Experiment	18	73,00	82,00	76,5000	2,33263
Control	25	70,00	80,00	73,6000	3,52373
Total	43	70,00	82,00	74,8140	3,37537

Normality

One-Sample Kolmogorov-Smirnov Test

		posttest_result
N		43
Normal Parameters ^{a,b}	Mean	74,8140
	Std. Deviation	3,37537
	Absolute	,129
Most Extreme Differences	Positive	,129
	Negative	-,101
Kolmogorov-Smirnov Z		,847
Asymp. Sig. (2-tailed)		,470

Test of Homogeneity of Variances

posttest_result

Levene Statistic	df1	df2	Sig.
3,034	1	41	,089

T-Test

Group Statistics

	group_class	N	Mean	Std. Deviation	Std. Error Mean
posttest_result	Experiment	18	76,5000	2,33263	,54981
	Control	25	73,6000	3,52373	,70475

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
posttest_result	Equal variances assumed	3,034	,089	3,040	41	,004	2,90000	,95400	,97337	4,82663
	Equal variances not assumed			3,244	40,779	,002	2,90000	,89384	1,09455	4,70545



UNIVERSITAS MERDEKA MALANG PROGRAM STUDI D3 BAHASA INGGRIS

Kampus : Jl. Terusan Halimun No. 11B Malang - 65146 Telp. (0341) 568395 Ext. 772
<http://www.unmer.ac.id>

SURAT KETERANGAN

No: Ket - 01 /DIII-BI/UM/VIII/2018

Yang bertanda tangan di bawah ini:

Nama : Drs. Suatmo Panca Putra, M.Pd.
Pangkat/Gol : Asisten Ahli Madya, III.A
Jabatan : Ketua Program Studi D-III Bahasa Inggris
Satuan Kerja : Program Studi D-III Bahasa Inggris

Dengan ini menerangkan bahwa:

Nama : Musrina, S.Pd
Program Studi : S2 Bahasa Inggris
Universitas Asal : Universitas Muhammadiyah Malang

Yang bersangkutan telah melakukan penelitian pada bulan April s/d Juni 2018 dengan baik dan bertanggung jawab di Program Studi D-III Bahasa Inggris Universitas Merdeka Malang.

Demikian surat keterangan ini dibuat, agar dapat digunakan sebagaimana mestinya.

Malang, 9 Juli 2018

Ketua Program



Drs. ~~Suatmo Panca Putra~~, M.Pd.

Tesis Musrina (1)

ORIGINALITY REPORT

3%

SIMILARITY INDEX

3%

INTERNET SOURCES

0%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1

dspace.bilkent.edu.tr

Internet Source

3%

Exclude quotes

Exclude bibliography

Off

On

Exclude matches

< 3%

